AFRL-IF-RS-TR-2002-12
Final Technical Report
February 2002

SI-FI (SYNTHESIZING INFORMATION FROM FORENSIC INVESTIGATIONS)

WetStone Technologies, Inc.

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2002-12 has been reviewed and is approved for publication.

APPROVED: JOHN FELDMAN
Project Engineer

FOR THE DIRECTOR: WARREN H. DEBANY, Jr.
Technical Advisor
Information Grid Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFGB, 525 Brooks Road, Rome, NY 13441-4505. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Feb 02	Final Apr 99 - Aug 00

4. TITLE AND SUBTITLE

SI-FI (SYNTHESIZING INFORMATION FROM FORENSIC INVESTIGATIONS)

5. FUNDING NUMBERS

C - F30602-99-C-0011
PE - 61102F
PR - 2301
TA - 03
WU - 02

6. AUTHOR(S)

G. Hosmer, G. Gordon, C. Siedsma, J. Hosmer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

WetStone Technologies, Inc.
273 Ringwood Road
Freeville, NY 13068-9618

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFGB
525 Brooks Road
Rome, NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2002-12

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: John Feldman, IFGB, 315-330-2664

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This effort describes a trusted mechanism for sharing computer forensics evidence among forensic examiners and others associated with determination of causes for Cyber Space Events. The process employed is described, as well as the unique solution identified. A prototype employing unique concepts is described. The solution employs a powerful and secure architecture, functioning within an open framework, employing strong security features, and easily adaptable to new Cyber-Forensic Software Tools. A key element of the solution relies on a Cyber Space Analogy to evidence bags which are sealed and tamper resistant. The SI-FI prototype demonstrator represents an advance in Cyber Forensic Technology.

14. SUBJECT TERMS

Computer Forensics, Sharing Evidence, Distance Data Collaboration

15. NUMBER OF PAGES

28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

Table of Contents

Table of Figures

Executive Summary

Computer Forensics can be defined as "the discovery, analysis, and reconstruction of evidence extracted from computer systems, computer networks, computer media and computer peripherals, that allows us to answer the questions of Who, What, When, Where, How and Why". [1] Computer forensic specialists are responsible for digging into the bowels of a suspect system, and finding the secrets that lie within. These specialists, however, are hindered in doing their job for a number of reasons.

Modern computer systems are typically sold today with hard drive capacities 20 Gigabyte or greater. And, this applies to only a single computer system - when there is a network involved, the task of inspecting that much information is monumental. The data extraction process is tedious, and search capabilities are found in only a limited number of forensic tools available today. Also, visualization, classification and cross-reference capabilities are almost non-existent. A forensic examiner should be able to cross-reference information found in e-mail files and addresses, web visit lists, word processor files, contact lists, and any number of additional sources.

For at least three decades, the storage and retrieval of large data sets has been a highly innovative field of computer science research and development. [2] Yet, there still does not exist an integrated set of tools that provides adequate support for accessing, organizing, and annotating these data sets and the associated collections of experimental data and documentation. [3] Any newly developed tools must focus attention on the data only while hiding the complexity of the computational environment.

Our approach for this effort was to examine the current and future cyber-forensic technologies and to define and develop a synthesizing architecture and framework that will accommodate new cyber-forensic tools.

As a result, we developed a working knowledge of the features necessary to construct the layers (or levels of abstraction) that are necessary in order to effectively integrate and utilize computer forensic data. The focus of this effort was to research and identify promising techniques, evaluate current approaches, and define these layers and their interfaces. Finally, we designed and developed a prototype framework for sharing cyber forensic evidence.

The SI-FI prototype is a fully operational secure world-wide-web client server based application. The software for SI-FI has been developed in a heterogeneous fashion. All source code has been written in Java and Java Script for portability, and global accessibility.

Technical Approach

The following major tasks were completed during the SI-FI project.

Task 1: Cyber-Forensic Needs Analysis

During this task we examined a wide range of cyber-forensic technology, we examined and assessed the threats to information technology and network, and examined on-going research cyber-forensics, cryptography, intrusion detection, network forensics and trusted time stamping. The result of this task can be found in "Cyber Forensics 2000, First Annual Study of the State-of-the-Art In Cyber Forensics". [4]

Task 2: SI-FI Requirements Definition

During this task we defined the major requirements that a successful SI-FI architecture would need. We based this on the results of Task 1 and discussions with DoD, Law Enforcement, Business and Industry, and developers of commercial and government cyber-forensic technology developers.

Task 3: SI-FI Architecture Definition

Based on the requirements defined in Task 2 and fundamental computer science and information security principals we examined a variety of methods and approaches to architecting a framework for synthesizing information from forensic investigations. The resultant high-level architecture provides the basis for the SI-FI product.

Task 4: SI-FI Design

The design of SI-FI has been an interesting road. As technology continued to advance during the course of this effort, new approaches and techniques for meeting the requirement and realizing the basic architecture have evolved. We were able to advance the design and integrate the latest information technology and information security advancements into the design.

Task 5: SI-FI Prototype Implementation

This final task implemented a prototype proof-of-concept system. We believe that the resulting prototype realizes the goals of the SI-FI architecture and integrates the latest advancements in information technology and security into the prototype. The maturity of the prototype will be evaluated during the Cyber-Forensic Experiment, CFX-2000, in October 2000.

SI-FI Requirements

The following table summarizes the requirements that are addressed by the SI-FI system.

ID	Requirement
R0001	The SI-FI system must support a variety of evidence collection, extraction, examination and analysis technologies.
R0002	The SI-FI system must be heterogeneous in order to support evidence collection, extraction, examination and analysis technologies on multiple heterogeneous computing platforms. (i.e. Wintel, Linux, Solaris)
R0004	The approach must support a broad range of cyber-forensic data.
R0005	Digital evidence by its very nature can be collected globally; therefore, synthesis of collected evidence must be possible regardless where the physical systems or networks exist.
R0006	The examination or sharing of the collected evidence must also be globally accessible by experts, investigators, and examiners anywhere and at anytime.
R0007	The integrity of digital evidence must be maintained. The SI-FI system must protect the integrity of evidence throughout its useful life.
R0008	The privacy of digital evidence especially when communicated over public networks such as the internet must be maintained at all times.
R0009	Access to digital evidence must be regulated by a strict policy and only those with authenticated privileges should be granted access to specific evidence.
R0010	The digital evidence must be tamper-proof.
R0011	Any SI-FI system must be accessible and useable by a variety of users with different levels of technical knowledge. Whenever possible the handling, policies and procedures applied to digital evidence should mimic established evidence handling processes.
R0012	All derived evidentiary data should be referenced to the original source evidence.
R0013	Detailed audit trail information must be maintained regarding each piece of evidence collected. This can include information pertaining to how evidence was collected, by whom, when, how (what tools were used), investigative techniques employed, etc.

Figure #1: SI-FI Requirements

SI-FI Architecture

The diagram below depicts the overall conceptual view of the system.

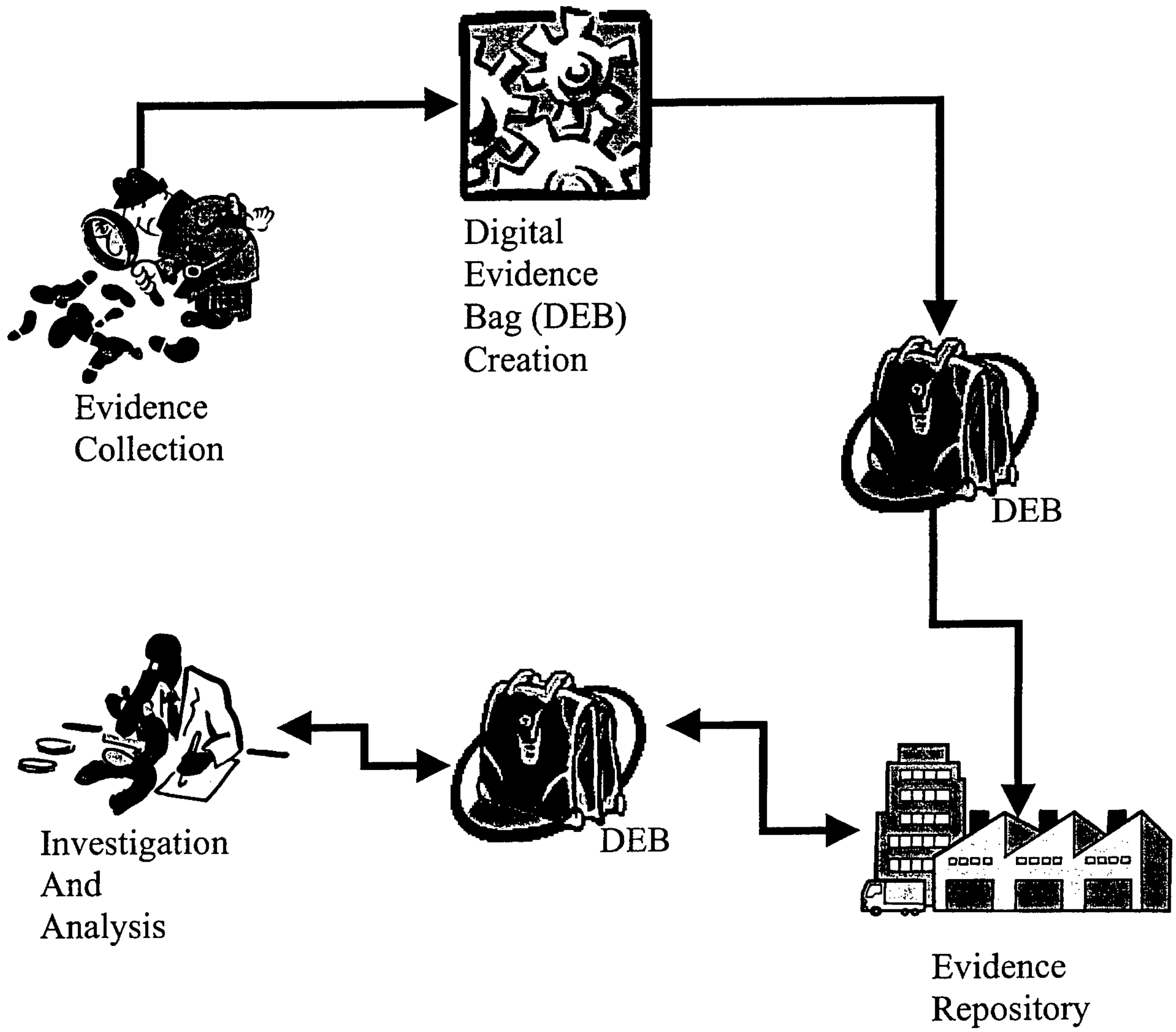

Figure #2: SI-FI Conceptual Overview

SI-FI is designed to support cyber-forensic collection, investigation and analysis processes. Existing tools and technologies in the form of commercial-off-the-shelf (COTS), government-off-the-shelf (GOTS), as well as, laboratory or research technologies perform one of these basic operations. These tools collect and extract digital evidence from a variety or sources or examine and analyze evidence thus collected. A conclusion reached after the examination of dozens of tools and technologies is that virtually all conform to one or both of those general paradigms.

We experimented with several design approaches to implementing SI-FI including traditional Application and Service Provider Interface constructs (API / SPI), traditional distributed database approaches, and finally with a completely heterogeneous web based design using modern development standards such as XML and Java. In the end we settled on the Web based XML / JAVA approach because it offered the greatest flexibility and robustness. The general SI-FI architecture is depicted below.

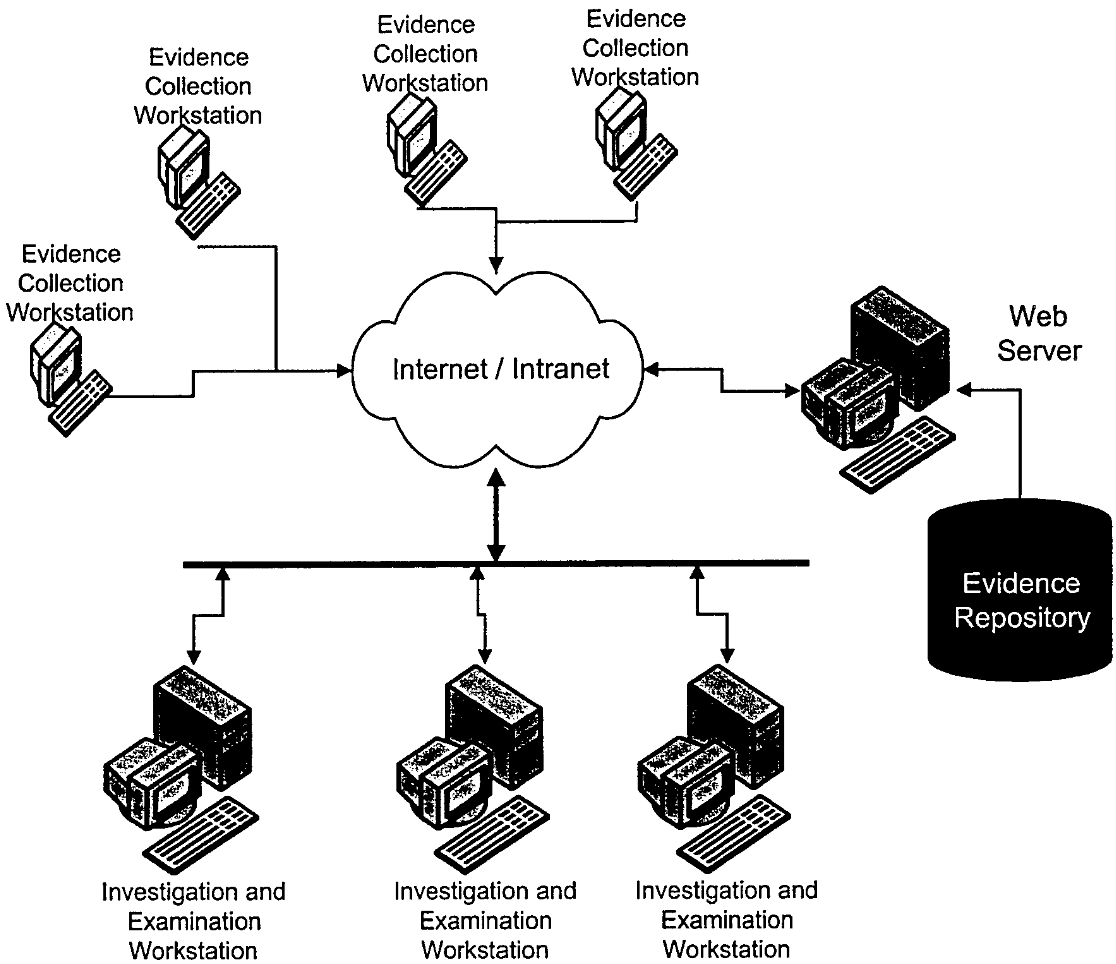

Figure #3: SI-FI Architecture

The SI-FI concept simple and comprises three basic tenants. First, creating a web-based evidence repository to hold and preserve digital evidence for on-going and archived cases. Secondly, providing the ability for Evidence Collection and storing at any location accessible to the Internet. Finally, allowing the retrieval of stored evidence from the repository for analysis and update as well as addition of new observations and hypotheses.

SI-FI Design

One problem that continues to hinder the advancement of cyber-forensic technologies is the definition of a data abstraction layer that will allow for the rapid innovation of evidence collection and investigation tools simultaneously, heterogeneously, and independently. In today's environment almost all cyber-forensic tools are point solutions that are self contained applications. While this approach has had limited success, it continues to be hindered by the lack of interoperability, difficulty in exchange of data among forensic tools and the difficulty of searching and correlating data from multiple cases.

We examined traditional evidence preservation methods such as the evidence bag shown below.

Figure #4: Traditional Evidence Bag

Traditionally, each bag of evidence is labeled using a widely accepted evidence marking system. The label information varies depending on the agency and the contents.

- Complete name (computer generated ID label-if available)
- Hospital number (computer generated ID label-if available)
- Date and time of collection
- Specimen type
- Location evidence was collected
- Signature of collector across the sealed bag

The examination of digital evidence can be treated similarly. Using the metaphor of the evidence bag the SI-FI design centers around the creation of digital evidence bags (DEBs). The concept is much like that of the physical evidence bag shown above, but instead of inserting physical devices such as weapons, blood swatches, knives, lifted foot or fingerprints the digital evidence bag contains digital data such as files, directories, slack space, deleted data, reports, electronic phone records, network traffic etc. Besides the evidence data the bag contains a label, as does the physical evidence bag. A variety of data can be stored in the DEB label. Unlike the contents of a physical evidence bag the data contained inside the digital evidence bag need not be physically removed in order to be examined. Instead a copy of the DEB contents can be removed and examined, analyzed and tested without disturbing the original contents of the bag. The DEB is tamper-proof.

DEB Design

The design of the DEB has taken on several intermediate forms along the way. We examined large and small databases, flat file information storage, information warehousing technologies and Internet -World Wide Web technologies. We selected the Extensible Markup Language or XML as the definition standard we would use to define the DEB.

In order to represent DEBs in SI-FI we have defined a Document Type Definition (DTD) in XML. Three separate XML representations were necessary to represent our SI-FI needs. These include the Digital Evidence Bag, Digital Evidence Descriptor, and the CASE. The next section defines the structured format for a digital evidence bag.

XML Representations for Digital Evidence

Digital Evidence Descriptor (DED)

Digital evidence is uniquely identified and preserved by ***binding*** information (such as ***who*** collected it, ***what*** was collected, ***why*** it was collected, ***where*** it was collected, and ***how*** it was collected) ***cryptographically*** with the *evidence file*. A Digital Evidence Descriptor (DED) XML schema was defined to facilitate storing that information and its cryptographic binding to the evidence file.

Digital Evidence Bag (DEB) schema

The Digital Evidence Bag (DEB) XML schema provides a container for defining the location of a DED and the location of the digital evidence file to which it is cryptographically bound. Digital Evidence can exist in multiple physical locations, yet still be the same piece of evidence. If the *physical location* of the evidence file and the actual *evidence file* were bound cryptographically the evidence would be bound to exist in a specific location. Organizations can share digital evidence by transferring DED(s)/evidence file(s) and store the evidence in new DEB(s).

Case (CASE)

The grouping of digital evidence at the highest level is accomplished by defining a Case (CASE) XML schema. A CASE contains information such as the identity of the person responsible for the case, legal information, and references (URLs) to all of the DEB(s) associated with the case. There is no cryptographic binding by the DED of evidence to a specific case.

The SI-FI prototype runs as a client/server product utilizing the Apache Web Server. SI-FI allows users to generate CASE(s), DED(s) and DEB(s) mimicking the conventional physical evidence paradigm used widely by law enforcement officials. DEB(s) and DED(s) must be associated with a specific case.

Important features of SI-FI include: Information Archiving, Evidence Preservation & Organization, Information Type, Semantic Identification, Evidence Mining, and Evidence Viewing Techniques. To provide these features many different technologies were employed. The design of SI-FI provides unique solutions for the following requirements:

R1. - DEB/DED Generation:
An application that provides a user interface for entering digital evidence and the generation of DEBs/DEDs.

R2. - DEB/DED Storage/Retrieval:
A DEB/DED repository and a standard for storing and retrieving DEB(s)/DED(s).

R3. -DEB/DED Content Viewing:
Presenting a DEB and DED document to the end user.

The selection of XML as the basis for SI-FI DEDs/DEBs lead to a Web based architecture utilizing a standard Web Browser and Server. The Web browser provides an ability to enter digital evidence and generate a DED. The Web server provides an ability to host the DED/DEB repository. The mechanisms for generating the DED from within a browser, transferring the DED from the browser to the Web server, creating a DEB and storing the DED are the key elements of the design.

Currently, Internet Explorer version 5.0 provides the best support for XML. Netscape's anticipated release of Navigator, version 6, will fully support XML. Java also provides support

for XML. The Apache XML APIs (Application Program Interface) provide a SAX (Simple API for XML) parser and a set of classes allowing the creation of DOM (Document Object Module) - compliant objects. These APIs provide the ability to plug in any XML standard. The generation of a DED requires cryptographic and trusted time services. These services can more easily be integrated using Signed Applets and JNI (Java Native Interface).

The following diagrams represent the high-level description of the DTD's designed and developed for SI-FI. For detailed descriptions of the DTD's see the Data Abstraction Design Document. [5]

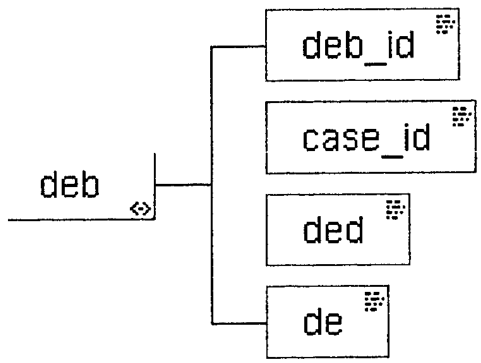

Figure #5: Digital Evidence Bag (DEB) – Document Type Definition (DTD)

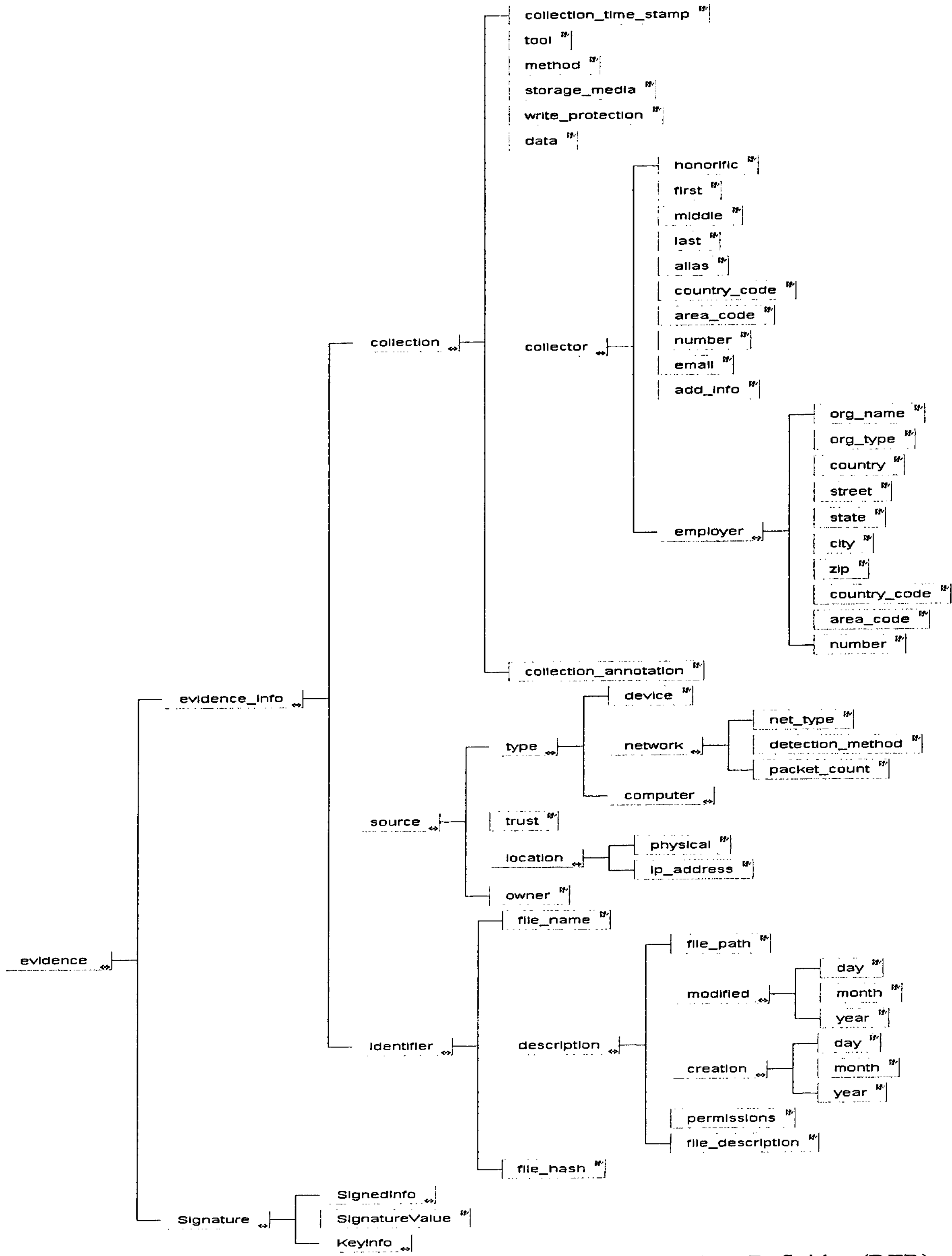

Figure #6: Digital Evidence Descriptor (DED) – Document Type Definition (DTD)

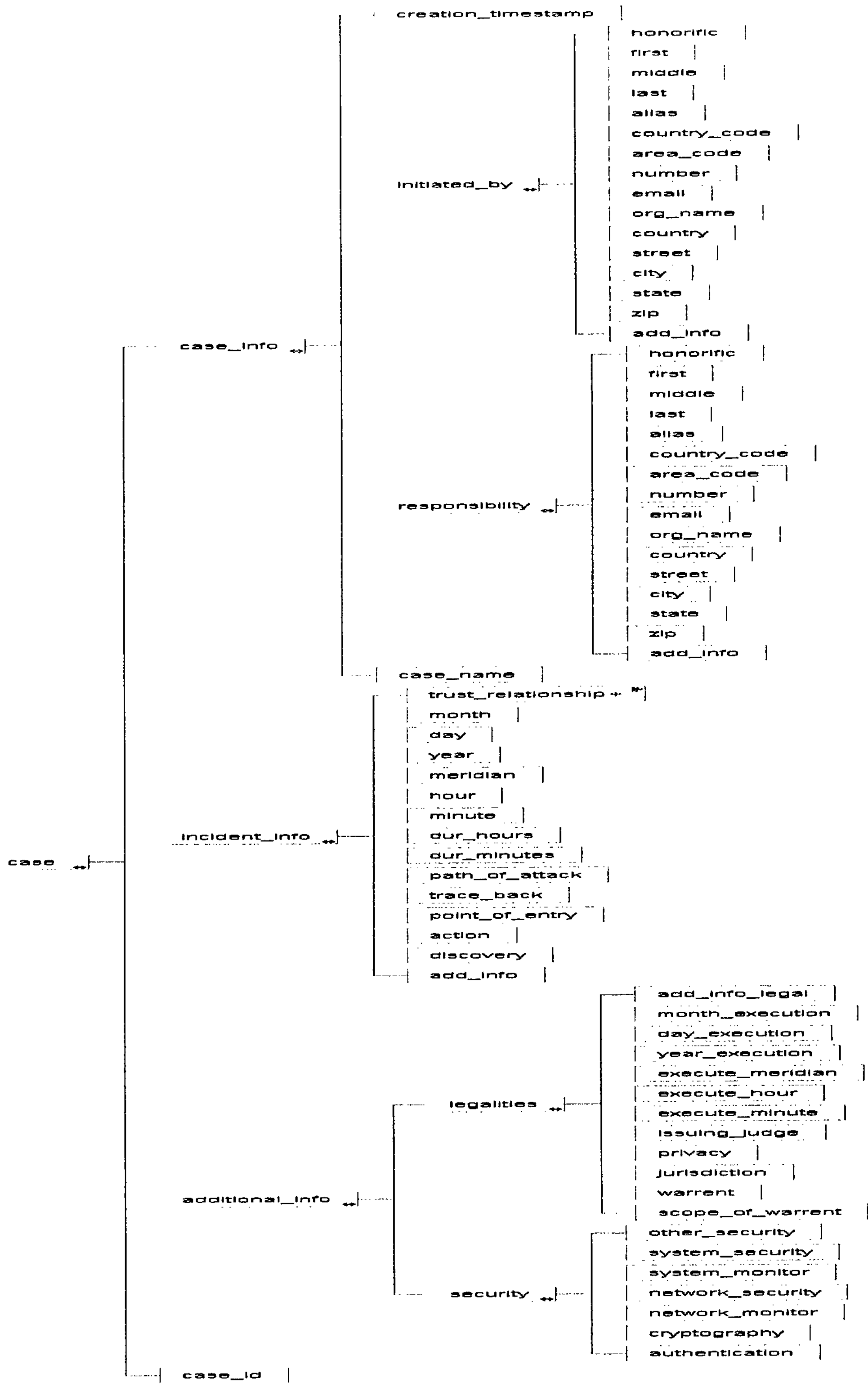

Figure #7: CASE– Document Type Definition (DTD)

Client Design Features

1. The main page for the SI-FI prototype contains an initial screen displaying information about the application and a frame used to invoke the different functions for creating a case, adding evidence, verifying evidence integrity, etc.

2. An HTML page for the input of information used to create a case. Information required by the CASE XML schema is entered here.

3. An HTML page for the input of information regarding a specific piece of evidence. The entry of the Digital Evidence Descriptor (DED) information is entered through a set of HTML pages containing form inputs that correspond to the DTD design. The form on each page represents a major node from within the DTD design. The forms that correspond to these nodes provide a one-to-one mapping of HTML form elements to the DTD node elements.

4. An evidence repository applet resides on the client machine. This applet is used for accessing the repository.

5. The evidence files and associated XML document are posted to the Web server via a secure process.

6. The main applet creates an XML document object as well as creating a hash-table for element lookup.

7. A key element involved in the saving of the XML document and associated files are the issue of security. Proper security settings are required to give applets security rights to the local files system and therefore a security policy must be implemented to ensure the access and integrity of all files stored on the local file system. A Java Plug-in environment is used for the interaction of the Applet with the local file system.

8. The transfer of the digital evidence descriptor (XML document and associated files) in a secure fashion is accomplished with the use of integrity and authentication technologies.

Server Design Features

1. The DED, consisting of the XML Document and the associated files will be authenticated and verified and stored on the server.

2. Users are identified and authenticated when they log into the server to retrieve information held within the digital evidence bags.

3. The digital evidence bags can be searched and organized.

Figure #8 presents a diagram of the SI-FI client-server architecture.

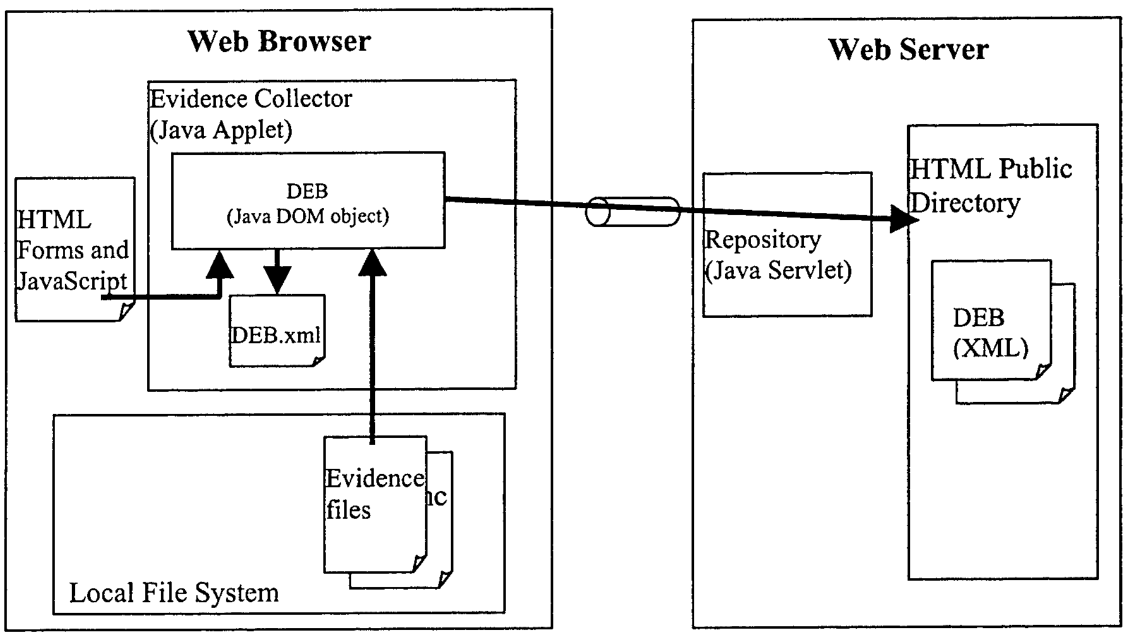

Figure #8: SI-FI Client-Server Architecture

Access Control

Users accessing the system are authenticated. The current system provides HTTP user authentication. This form of authentication requires a valid username and password to be submitted before any access to SI-FI resources are allowed. The fastest way to accomplish this is to use any web server. Apache Web Server will be used because it provides the ability to define users and restrict access to web resources based on user rights.

Repository Overview

The SIFI repository is responsible for storing, querying, and retrieving a case. A case is the composition of DEB(s), DED(s), and evidence file(s). The primary function of the repository is to provide access to a case via HTTP GET/POST methods. To keep the design and implementation simple all DEB(s), DED(s) and evidence file(s) are located in public directories so the web server can serve them without using a custom servlet. Since Apache Web Server provides user authentication and access control, no unauthorized users can have access to those public files.

Repository Storage

Storing the DEB(s), DED(s), and evidence file(s) requires a custom servlet. This servlet will interact with an applet running within the context of the users web browser. A unique case identifier and a DEB identifier that is unique within a case are required when storing case data. When a user creates/updates a case with new evidence the applet will send the case id, along with associated DEBs and evidence files to the servlet. The servlet will store each DED and evidence file on the server.

Repository Querying

The client is presented with a table of contents that displays a listing of the cases, and the corresponding DEB(s) that comprise that particular case. All entries are in the form of an http reference to the actual document on the server.

Repository Retrieval

All documents exist in a public directory structure, which can be served by the Web Server. The web pages displayed via querying the repository contain http references to the DEB(s) and evidence files. The client's browser can download any DEB or evidence file using standard browser services. This page also contains an applet that can verify the integrity of the evidence file(s) based on the information stored in the DEB.

SI-FI Prototype

The fully operational prototype can be accessed at www.wetstonelabs.com. A user ID and password are needed to access and experiment with the capabilities. WetStone Technologies and AFRL/IFGB must be contacted in order to provide access to the prototype system.

The following figures are screen shots from the operational prototype.

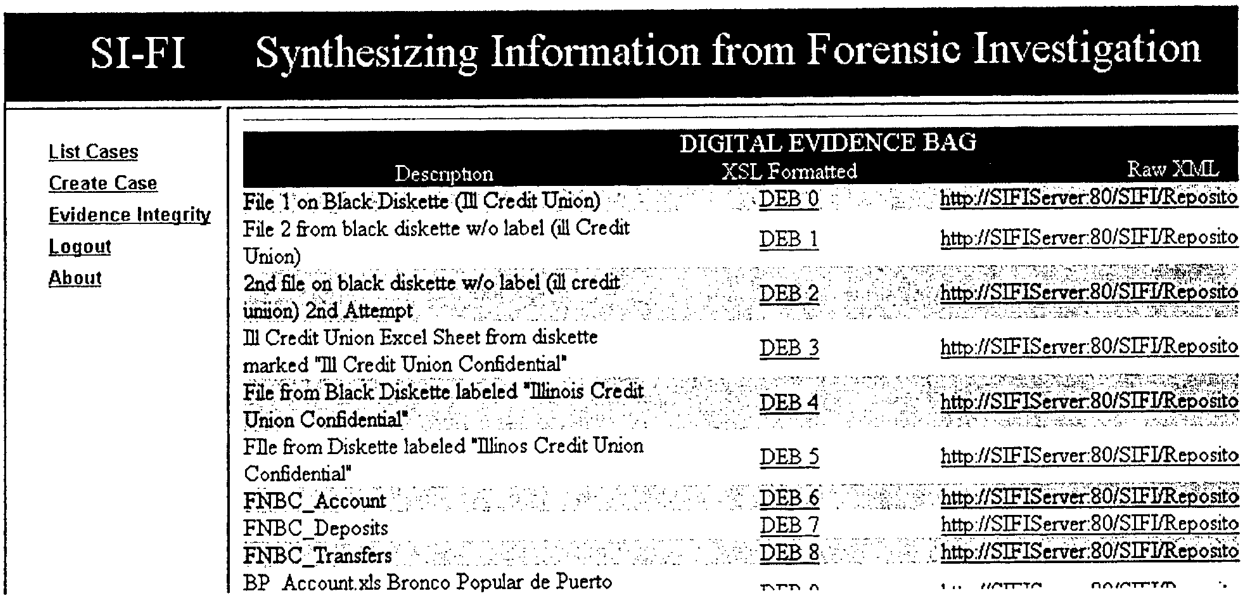

Figure # 9: DEB- Digital Evidence Bag

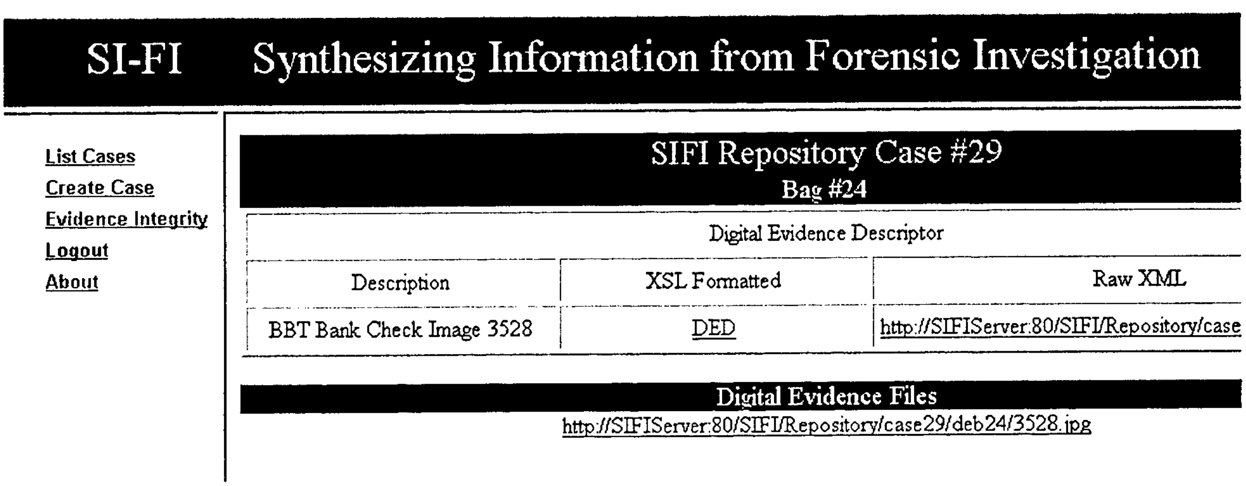

Figure # 10: Evidence Repository

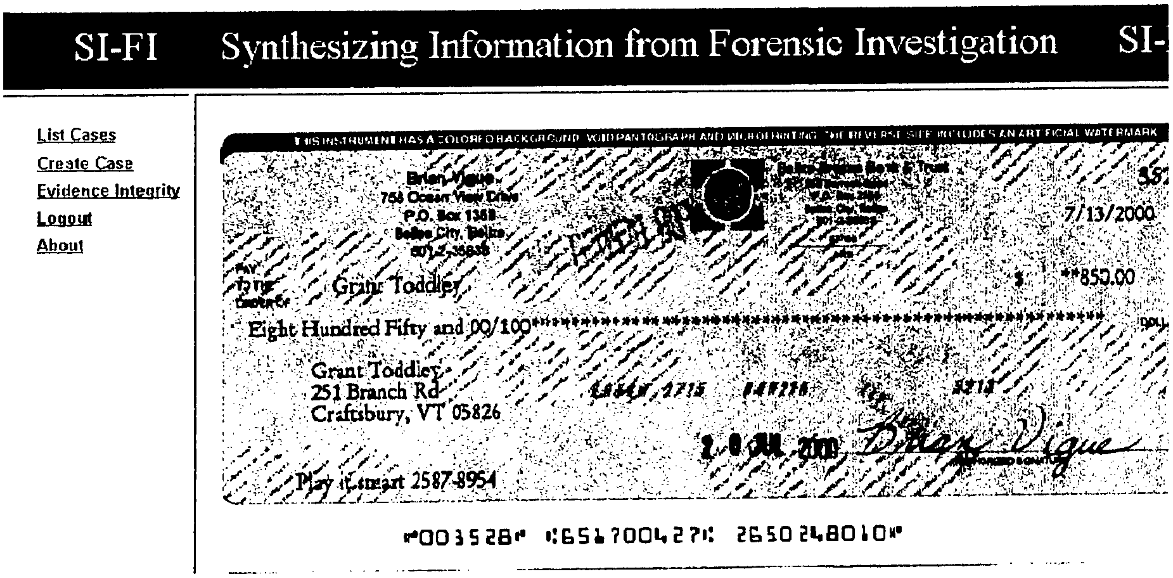

Figure #11: Sample Evidence

Conclusions

During this effort, WetStone Technologies with the assistance of Dr. Gary Gordon has taken the concept of the synthesis of information from forensic investigations from concept to operational prototype. The major objectives of the effort have been exceeded. The SI-FI prototype should serve as a foundation for additional research, development and experimentation in the area of cyber-forensics.

This project has accomplished fundamental tasks and developed innovative tools for use in cyber forensic investigation. Research into new methods and approaches have advanced the concept of sharing and synthesizing information from forensic investigations. Requirements for the synthesis of such forensic data have been established and documented. The first conceptual and functional Digital Evidence Bag for the archival and preservation of digital evidence has been created. The first conceptual and functional distributed architecture and framework for the secure global sharing of cyber-forensic data has been created. An operational prototype of the framework has been created and has been used as the underlying framework for the support of the Cyber-Forensic Experiment (CFX-2000) in October, 2000. As mentioned earlier in this document, the first Annual Cyber-Forensics Study was also been created as a result of this project.

References

[1] C. Hosmer, G. Gordon, "Advancing Crime Scene Computer Forensic Techniques", SPIE Conference, November 98.
[2] C.J. Breiteneder, M. Hitz & T.A. Mueck, "Metadata Mining in Legacy Data Sets". http://computer.org/conferen/meta96/breitender/IEEE.fmk.html
[3] R. Springmayer, N. Werner, J. Long, "Mining Scientific Data Archives through Metadata Generation". Proceedings of the First IEEE Metadata Conference, April, 1996.
[4] C. Hosmer, G. Gordon, C. Hyde, T. Grant, "Cyber Forensics 2000 – 1st Annual Study of the State-of-the-Art in Cyber Forensics, August 2000.
[5] C. Hosmer, C. Siedsma, J. Hosmer, G. Gordon, "SI-FI Data Abstraction Layer Design Document", August, 2000.

MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)

The advancement and application of Information Systems Science and Technology to meet Air Force unique requirements for Information Dominance and its transition to aerospace systems to meet Air Force needs.